

**PENINGKATAN KUALITAS SINYAL SUARA
DENGAN EKSPANSI LEBAR PITA SECARA ARTIFISIAL**

Riko Arlando Saragih¹

ABSTRACT

In most communication systems, speech was transmitted in narrowband frequency, mainly at frequency 300 Hz to 3400 Hz. Therefore, the frequencies outside this narrow bandwidth were lost. The lost frequencies can cause bias articulation in some speech.

Bandwidth expansion is one technique to overcome this problem. Many bandwidth expansion techniques have been developed to increase the speech signal bandwidth on communication channels through envelope aliasing, nonlinear processing, and linear model of human speech processing.

Artificial Bandwidth Expansion (ABE) is one of the bandwidth expansion techniques to increase speech signal sampling frequency (upsampling) from 8 kHz to 16 kHz. The process annexes extra spectral components which are the images of the low frequency spectral components (narrowband) in the higher frequency segment. This method works frame by frame and each frame is 30 ms long and 5 ms overlapping with the adjacent frames. In this research, every frame was classified using Hidden Markov Model (HMM) and stated as sibilants or fricative consonants or stop consonants and treated separately (attenuated or amplified) according to the classification results.

Comparison tests were done in this research based on the audio perceptions of some responders. Based on the results, it was acknowledged that the quality of the speech signal was significantly improved although some distortions were somewhat objectionable.

Keywords : artificial bandwidth expansion, speech, HMM, classification

INTISARI

Dalam sistem komunikasi yang ada, sinyal suara manusia dikirim di dalam pita frekuensi yang sempit, yaitu sepanjang frekuensi 300 Hz sampai 3400 Hz. Hal tersebut mengakibatkan frekuensi di luar lebar pita tersebut menjadi hilang. Hilangnya frekuensi tersebut mengakibatkan artikulasi pada beberapa ucapan menjadi kurang jelas.

Ekspansi lebar pita adalah suatu teknik untuk mengatasi masalah ini. Banyak metode ekspansi lebar pita yang telah dikembangkan untuk meningkatkan lebar pita frekuensi percakapan pada kanal komunikasi, seperti metode aliasing selubung, pemrosesan secara nonlinear, dan model linear untuk proses produksi sinyal suara.

Ekspansi lebar pita secara artifisial adalah salah satu teknik ekspansi lebar pita yang bekerja dengan cara menaikkan frekuensi sampling (*upsampling*) sinyal suara dari 8 kHz menjadi 16 kHz. Proses tersebut mengakibatkan komponen spektral yang merupakan cerminan komponen spektral frekuensi rendah muncul pada frekuensi tinggi. Metode ini bekerja *frame per frame* dan setiap *frame* memiliki panjang 30 ms serta memiliki *overlapping* selama 5 ms dengan *frame* yang bersebelahan. Dalam penelitian ini setiap *frame* diklasifikasikan dengan menggunakan *Hidden Markov Model* (HMM) dan dinyatakan sebagai *frame* yang berbunyi desis atau konsonan frikatif atau konsonan stop dan diproses secara berbeda (diredam atau dikuatkan) sesuai dengan hasil klasifikasi

Pengujian yang dilakukan pada penelitian ini adalah dengan cara memperdengarkan sinyal suara kepada beberapa pendengar. Berdasarkan hasil uji, diketahui bahwa kualitas sinyal suara dapat diperbaiki walaupun distorsi masih terdengar.

Kata kunci : ekspansi lebar pita secara artifisial, sinyal suara, HMM, klasifikasi

PENDAHULUAN

Dalam komunikasi tradisional seperti pada jaringan PSTN atau GSM, sinyal suara memiliki lebar pita sekitar 300

Hz – 3400 Hz dan biasa disebut sebagai sinyal pita sempit (*narrowband*). Hal tersebut dapat mengakibatkan spektrum si-

¹ Jurusan Teknik Elektro UKM, riko_saragih@yahoo.com

nyal suara dengan frekuensi di luar lebar pita tersebut menjadi hilang. Hilangnya spektrum frekuensi ini mengakibatkan artikulasi pada beberapa ucapan menjadi kurang jelas.

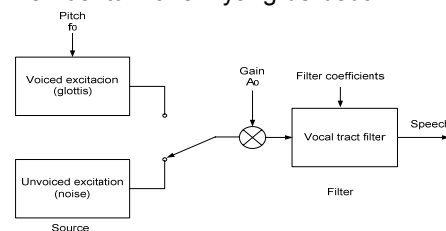
Beberapa usaha telah dilakukan untuk menemukan suatu metode yang dapat meningkatkan lebar pita frekuensi percakapan yang dapat mengakomodasi frekuensi tinggi pada suara manusia agar sinyal suara yang ditransmisikan memiliki kualitas yang cukup tinggi. *Artificial Bandwidth Expansion (ABE)* adalah suatu metode ekspansi lebar pita secara artifisial yang dapat mengubah sinyal pita sempit menjadi sinyal pita lebar (*wideband*). Untuk melakukan ekspansi lebar pita, ABE menggunakan teknik *up-sampling* frekuensi sampling yang dilanjutkan dengan modifikasi spektrum pada bunyi yang memiliki artikulasi kurang jelas.

Selain ABE, terdapat beberapa metode yang sering digunakan untuk ekspansi frekuensi, seperti aliasing selubung atau *envelope aliasing* (Yasukawa, 1994), pemrosesan secara nonlinear (Yasukawa, 1996), dan model linear untuk sistem produksi sinyal suara pada manusia (Kallio, 2002).

Pada ekspansi frekuensi yang menggunakan metode aliasing selubung (Yasukawa, 1994), sinyal pita sempit ($f_{\text{sampling}} = 8 \text{ kHz}$) di-*upsampling*, sehingga frekuensi yang sebelumnya mengalami aliasing dapat muncul pada frekuensi tinggi. Frekuensi sampling diubah menjadi 14 kHz dan cerminan sinyal pita sempit dapat terlihat pada daerah frekuensi 3,5 kHz – 7 kHz. Sinyal suara ini kemudian difilter dengan menggunakan filter *high pass* dengan frekuensi *cut-off* 3,5 kHz. Komponen spektral dibentuk melalui suatu filter pembentuk (*shaping filter*) yang berupa filter *low-pass Finite Impulse Response (FIR)* fase linear dengan frekuensi *cut-off* 2 kHz dan frekuensi *stopband* 8 kHz. Energi sinyal yang terbentuk disesuaikan dengan sinyal pita sempit semula. Pengaturan level diperlukan untuk mengatur level energi frekuensi tinggi ini pada level yang sesuai. Sebagai hasilnya sinyal pita lebar yang diperoleh merupakan hasil penjumlahan atas sinyal frekuensi tinggi yang terbentuk dan sinyal pita sempit asli.

Sedangkan pada pemrosesan secara nonlinear (Yasukawa, 1996), prosesnya hampir sama dengan aliasing selubung, tetapi pada metode ini komponen spektral dihasilkan dengan melakukan penyearahan (*rectification*), sehingga harmonik dari sinyal asli muncul pada frekuensi tinggi. Komponen spektral dimodifikasi menggunakan filter pembentuk dengan karakteristik yang sama pada metode aliasing selubung. Frekuensi tinggi yang ditambahkan diatur pada level yang sesuai dan sinyal pita lebar dihasilkan dari penjumlahan sinyal frekuensi tinggi ini dan sinyal pita sempit asli.

Ekspansi lebar pita juga dapat dilakukan dengan menggunakan model linear sistem produksi sinyal suara pada manusia (Kallio, 2002). Pada model tersebut, sinyal eksitasi difilter menggunakan sebuah filter yang berisi informasi *vocal tract*. Terdapat dua sinyal eksitasi, yaitu suara *voiced* dan *unvoiced*. Suara *voiced* terjadi pada pulsa *glottal* yang kemudian mengalami turbulensi aliran udara dari *vocal tract*. *Vocal tract* mengubah frekuensi sinyal eksitasi ini dengan cara mengubah bentuk saat ucapan membentuk fonem yang berbeda.



Gambar 1. Model Sistem Produksi Suara Manusia

Gambar 1 memperlihatkan proses pembangkitan sinyal suara manusia untuk frekuensi tinggi dengan menggunakan model *source-filter*. Sinyal eksitasi menggambarkan struktur yang baik dari spektrum dan digunakan sebagai informasi dari sinyal pita sempit. Sinyal tersebut merupakan Sinyal residu di-*upsampling* dua kali dari frekuensi sampling dan frekuensi aliasing difilter dengan filter *low-pass*. Kemudian hasil proses *upsampling* difilter dengan filter sintesis yang memprediksi selubung pita-lebar. Sinyal suara hasil ekspansi ini merupakan kombinasi atas sinyal pita sempit asli dan

sinyal frekuensi tinggi dari keluaran filter sintesis.

Sinyal suara manusia merupakan sinyal yang berubah terhadap waktu dengan kecepatan yang relatif rendah. Jika diamati pada selang waktu yang pendek (antara 5 ms sampai 100 ms), karakteristiknya bersifat tetap, tetapi jika diamati pada selang waktu yang lebih lama akan terlihat karakteristik yang berubah-ubah sesuai dengan kalimat yang diucapkan.

Terdapat berbagai cara untuk mengklasifikasikan bagian-bagian atau komponen sinyal suara. Salah satu cara yang sederhana adalah dengan mengklasifikasikan bagian-bagian sinyal suara menjadi tiga bagian, yaitu : (Rabiner dan Juang, 1993)

(1) *Unvoiced*

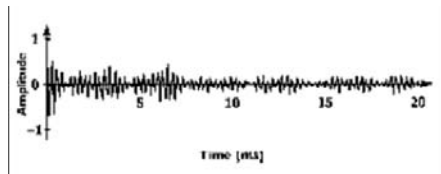
Keadaan pada saat *vocal cord* tidak melakukan vibrasi sehingga suara yang dihasilkan bersifat tidak periodis atau bersifat acak.

(2) *Voiced*

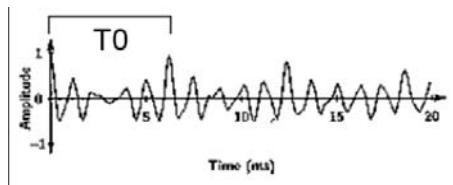
Keadaan pada saat terjadinya vibrasi pada *vocal cord* sehingga menghasilkan suara yang bersifat kuasi periodis.

(3) *Silence*

Keadaan pada saat tidak ada sinyal suara yang diucapkan.



Gambar 2. Sinyal Suara *Unvoiced*



Keterangan : T0 adalah periode *pitch*.

Gambar 3. Sinyal Suara *Voiced*

Unit terkecil sinyal suara yang dapat dibedakan manusia disebut fonem

(Rabiner dan Juang, 1993). Himpunan fonem yang ada dalam suatu bahasa berbeda-beda.

Setiap fonem dinyatakan dengan suatu simbol yang unik. Terdapat beberapa standar penamaan fonem, seperti IPA (*International Phonetic Alphabet*), ARPABET, serta SAMPA.

Beberapa jenis fonem dalam Bahasa Inggris seperti berikut: (Rabiner dan Juang, 1993)

(1) Vokal

Setiap vokal memiliki komponen frekuensi tertentu yang membedakan karakter satu fonem vokal dengan fonem vokal lainnya. Fonem vokal Bahasa Inggris mencakup fonem-fonem /IY/, /IH/, /EH/, /AE/, /AA/, /ER/, /AH/, /AX/, /AO/, /UW/, /UH/, dan /OW/.

(2) Diftong

Diftong pada prinsipnya adalah dua fonem vokal yang berurutan dan diucapkan tanpa jeda. Fonem diftong Bahasa Inggris mencakup /AY/, /OY/, /AW/, dan /EY/. Karakteristik diftong mirip dengan karakteristik fonem-fonem vokal pembentuknya disertai bentuk transisinya.

(3) Konsonan Nasal

Konsonan nasal dibangkitkan dengan eksitasi *glottal* dan *vocal tract* yang mengkerut total pada beberapa titik tertentu sepanjang lintasan pengucapan. Bagian belakang langit-langit merendah, sehingga udara mengalir melalui *nasal tract* dengan suara yang dipancarkan melalui lubang hidung.

(4) Konsonan Frikatif

Konsonan frikatif pada prinsipnya dibedakan menjadi frikatif *unvoiced* dan *voiced*. Frikatif *unvoiced* dibentuk dengan suatu eksitasi terhadap *vocal tract* dengan suatu aliran udara yang tetap, sehingga menyebabkan turbulensi di daerah yang mengkerut dalam *vocal tract*. Frikatif *voiced* sedikit berbeda dengan frikatif *unvoiced*. Pada frikatif *voiced*, suara dihasilkan oleh dua sumber eksitasi, yaitu *vocal tract* dan *glottis*. Fonem Bahasa Inggris yang termasuk frikatif *unvoiced* adalah /F/, /TH/, /S/, dan

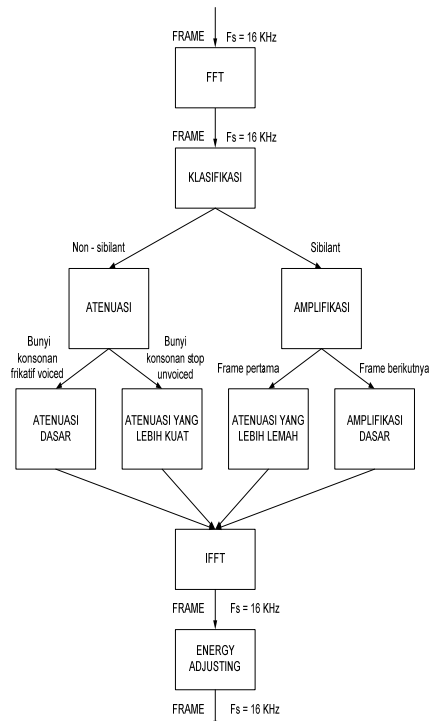
/SH/, sedangkan yang termasuk frikatif *voiced* adalah /V/, /Z/, dan /ZH/.

(5) Konsonan Stop
Konsonan stop dibedakan menjadi konsonan stop *unvoiced* dan *voiced*. Konsonan stop memiliki bentuk yang berbeda dengan konsonan lainnya. Konsonan ini memperlihatkan pola transien dan tidak kontinu. Konsonan ini dibentuk dengan cara memberikan tekanan pada kondisi penge-rutan total di bagian rongga mulut tertentu dan segera diikuti dengan pelemasan. Selama periode penge-rutan total terjadi, tidak ada suara yang dikeluarkan dari mulut, sehingga fonem ini selalu mengandung bagian yang menyerupai *silence*. Fonem Bahasa Inggris yang termasuk konsonan stop *unvoiced* adalah /P/, /T/, dan /K/, sedangkan yang termasuk konsonan stop *voiced* adalah /B/, /D/, dan /G/.

Gambar 4 memperlihatkan dia-gram blok proses ekspansi lebar pita secara artifisial pada sinyal suara yang dilakukan pada penelitian ini. Pada dasarnya proses ini terdiri dari lima tahapan utama, yaitu *framing*, *up-sampling*, kla-sifikasi, amplifikasi/atenuasi, dan penyesuaian energi (*energy adjusting*).

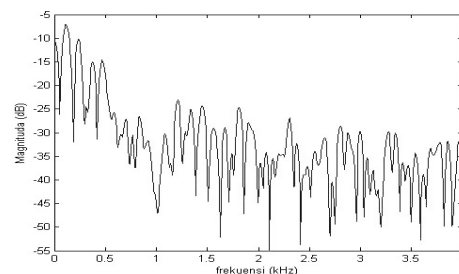
Pada tahap *framing*, sinyal suara masukan dibagi menjadi beberapa *frame*. Setiap *frame* memiliki lebar 20 ms, dengan frekuensi sampling sebesar 8 kHz, sehingga terdapat 160 sampel pada setiap *frame*. Untuk membagi sinyal masukan menjadi *frame* yang lebih ke-cil, dilakukan proses *windowing* (penjen-dela-an) dengan lebar 30 ms (240 sampel) dengan *window* Hamming. Hal ini mengakibatkan pada setiap *frame* me-ngalami *overlapping* selama 5 ms (40 sampel) pada kedua sisi dengan *frame* tetangga (yang bersebelahan).

Proses *upsampling* bertujuan untuk menaikkan frekuensi sampling dari sinyal suara. Pada penelitian ini, setiap *frame* mengalami *upsampling* sebesar dua kali frekuensi sampling semula, ya-itu dari 8 kHz menjadi 16 kHz.

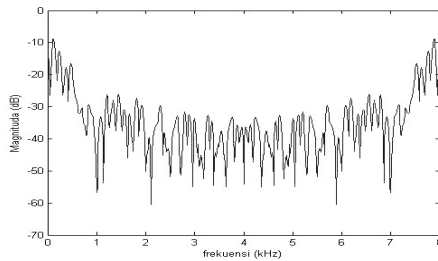


Gambar 4. Diagram Blok Ekspansi Frekuensi secara Artifisial

Proses *upsampling* yang dimaksud adalah menyisipkan nol pada tiap sampel. Akibat proses ini, frekuensi harmonik pada sinyal pita sempit terlihat pada spektrum di daerah 4 kHz – 8 kHz pada sinyal pita lebar. Gambar 5 dan Gambar 6 berturut-turut memperlihatkan contoh spektrum sinyal suara sebelum dan sesudah di-*upsampling*.



Gambar 5. Contoh Spektrum Sinyal Suara Sebelum Di-*upsampling*



Gambar 6. Contoh Spektrum Sinyal Suara Sesudah Di-*upsampling*

Komponen spektral dari *frame* dimodifikasi dengan fungsi amplifikasi atau atenuasi, tetapi tidak semua *frame* mengalami modifikasi, melainkan hanya *frame* yang berbunyi desis (*sibilant*), konsonan frikatif *voiced* dan konsonan stop *unvoiced* saja yang mengalami modifikasi. Hal ini disebabkan oleh karena ketidakjelasan artikulasi pada bunyi-bunyi tersebut. Untuk itu dilakukan klasifikasi pada tiap *frame* dan dalam penelitian ini klasifikasi dilakukan dengan menggunakan *Hidden Markov Model* (HMM).

HMM merupakan pengembangan atas *Markov Chain*, yaitu fungsi peluangnya menggambarkan simbol keluaran yang merupakan gabungan dari *state* dan perpindahan *state*. HMM sebagai simbol observasi diskret memiliki elemen-elemen sebagai berikut (Rabiner, 1989) :

- (1) Jumlah *state* dalam HMM adalah N . Individu *state* diberi nama $\{1, 2, \dots, N\}$ dan *state* pada waktu t adalah q_t .
- (2) Jumlah simbol observasi yang terlihat pada setiap *state* adalah M . Simbol observasi merupakan keluaran fisik dari sistem yang dimodelkan. Individu simbol dilambangkan dengan $V = \{v_1, v_2, \dots, v_M\}$.
- (3) Distribusi peluang transisi *state* $A = \{a_{ij}\}$, dimana :

$$a_{ij} = P[q_{t+1} = j \mid q_t = i], \quad (1)$$
 dengan $1 \leq i, j \leq N$.

Jika sebuah *state* dapat dicapai dari *state* lain dalam satu langkah, maka $a_{ij} > 0$ untuk semua i dan j .

- (4) Distribusi peluang simbol observasi $B = \{b_j(k)\}$, dimana :

$$b_j(k) = P[o_t = v_k \mid q_t = j], \quad (2)$$
 dengan $1 \leq k \leq M$ dan $b_j(k)$ adalah distribusi simbol pada *state* j , $j = 1, 2, \dots, N$.

- (5) Distribusi peluang observasi $\pi = \{\pi_j\}$, dimana :

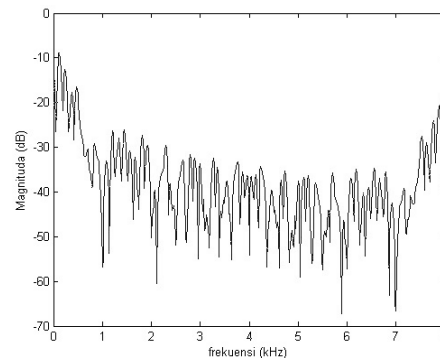
$$\pi_i = [q_t = i], \quad 1 \leq i \leq N \quad (3)$$

Keterangan : HMM dapat dinotasikan sebagai model $\lambda = (A, B, \pi)$.

Salah satu cara klasifikasi dengan HMM adalah dengan memanfaatkan struktur matriks transisinya. Dalam penelitian ini parameter model HMM tiap fonem yang diestimasi adalah parameter-parameter jumlah *hidden state*, *left-right topology*, dan *covariance*. Untuk mendapatkan model HMM untuk setiap fonem yang perlu diklasifikasikan, maka dilakukan ekstraksi ciri pada tiap *frame* menggunakan frekuensi *formant*.

Bunyi konsonan frikatif *voiced* pada frekuensi tinggi memiliki amplitude yang lebih rendah, sehingga pada daerah frekuensi tinggi komponen spektral harus diredam untuk mendapatkan bunyi yang alami. Pada konsonan stop *unvoiced* juga mengalami hal yang sama.

Fungsi atenuasi yang diberikan untuk konsonan frikatif *voiced* adalah atenuasi 20 dB/dekade untuk frekuensi 4 kHz sampai dengan 8 kHz. Untuk konsonan stop *unvoiced* fungsi atenuasi yang diberikan lebih kuat, yaitu 40 dB/dekade untuk frekuensi 4 kHz sampai dengan 8 kHz. Gambar 7 memperlihatkan contoh spektrum sinyal suara setelah mengalami atenuasi.

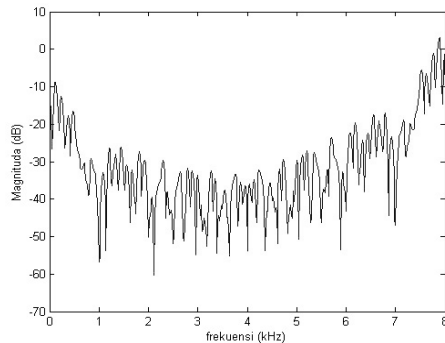


Gambar 7. Contoh Spektrum Sinyal Suara Setelah Mengalami Atenuasi

Pada bunyi *sibilant*, amplitude pada frekuensi tinggi lebih tinggi, sehingga untuk memperbaiki komponen spektral pada bunyi *sibilant* tersebut perlu dilakukan suatu fungsi amplifikasi. Untuk bunyi *sibilant* pada urutan pertama

diberikan fungsi penguatan sebesar 10 dB/dekade untuk frekuensi 4 kHz sampai dengan 8 kHz. Hal ini untuk memperhalus transisi dari *frame* sebelumnya. Untuk bunyi *sibilant* pada *frame* berikutnya diberikan fungsi penguatan sebesar 20 dB/dekade untuk frekuensi 4 kHz sampai dengan 8 kHz.

Karena struktur harmonik pada pita sempit muncul pada daerah frekuensi tinggi pita lebar, maka untuk semua *frame*, komponen spektral frekuensi tinggi pita lebar diperhalus. Gambar 8 memperlihatkan contoh spektrum sinyal suara setelah mengalami amplifikasi.



Gambar 8. Contoh Spektrum Sinyal Suara Setelah Mengalami Amplifikasi

Setelah dilakukan modifikasi komponen spektral pada domain frekuensi, maka untuk mengembalikan *frame* pada domain waktu dilakukan *Inverse Fast Fourier Transform* (IFFT) sebanyak 1024 titik untuk tiap *frame*. IFFT ini menghasilkan 480 sampel (30 ms pada $f_s = 16$ kHz) untuk tiap *frame*.

Karena terjadi penambahan komponen spektral, maka perlu dilakukan penyesuaian energi. Penyesuaian energi dilakukan dengan memberikan gain sebesar 2 untuk bagian sinyal yang mengalami ekspansi lebar pita secara artifisial. Hal tersebut dilakukan karena sinyal mengalami *upsampling* sebesar dua kali frekuensi sampling sebelumnya.

PEMBAHASAN

Sinyal suara masukan dalam penelitian ini berupa sinyal suara yang disimpan dalam *file* dengan ekstensi *wav* dengan format digital PCM (*Pulse Code Modulation*) dengan frekuensi sampling

8000 Hz dengan resolusi 16 bit (mono). Sinyal suara yang akan diujikan berupa 7 sinyal suara yang diucapkan oleh 2 orang wanita dan 1 pria yang direkam terpisah melalui program *sound recorder* pada *personal computer* yang dilengkapi *microphone*. Tiap sinyal suara berisi ucapan yang mengandung fonem yang mungkin akan mengalami modifikasi.

Model HMM untuk setiap fonem yang perlu diklasifikasikan pada penelitian ini memiliki 3 *state*, yaitu setiap *state* memiliki 120 observasi.

Pada proses pelatihan didapatkan matriks transisi *state* (A) dan matriks peluang observasi (π). Matriks model HMM yang diperoleh dari pelatihan adalah sebagai berikut :

- (1) Untuk *state* sekarang adalah konsonan frikatif dengan *state* berikutnya adalah konsonan stop dan sibilant, maka matriks distribusi peluangnya (π) adalah $\begin{bmatrix} 0.83418 \\ 0.093413 \\ 0.07241 \end{bmatrix}$ dan matriks

distribusi peluang transisi *state*-nya (A) adalah

$$\begin{bmatrix} 0.31457 & 0.20339 & 0.48204 \\ 0.52735 & 0.12636 & 0.34629 \\ 0.22175 & 0.34918 & 0.42907 \end{bmatrix}$$

- (2) Untuk *state* sekarang adalah konsonan frikatif dengan *state* berikutnya adalah *sibilant* dan konsonan stop, maka matriks distribusi peluangnya (π) adalah $\begin{bmatrix} 0.85872 \\ 0.11001 \\ 0.03126 \end{bmatrix}$ dan matriks

distribusi peluang transisi *state*-nya (A) adalah

$$\begin{bmatrix} 0.21952 & 0.12336 & 0.65712 \\ 0.047612 & 0.27875 & 0.67364 \\ 0.16824 & 0.62664 & 0.20512 \end{bmatrix}$$

- (3) Untuk *state* sekarang adalah konsonan stop dengan *state* berikutnya adalah *sibilant* dan konsonan frikatif, maka matriks distribusi peluangnya (π) adalah $\begin{bmatrix} 0.84033 \\ 0.15071 \\ 0.008962 \end{bmatrix}$ dan matriks

distribusi peluang transisi *state*-nya (A) adalah

$$\begin{bmatrix} 0.20454 & 0.5718 & 0.22366 \\ 0.21163 & 0.12247 & 0.6659 \\ 0.11067 & 0.086075 & 0.80326 \end{bmatrix}$$

- (4) Untuk *state* sekarang adalah konsonan stop dengan *state* berikutnya adalah konsonan frikatif dan *sibilant*, maka matriks distribusi peluangnya

$$(\pi) \text{ adalah } \begin{bmatrix} 0.81896 \\ 0.059319 \\ 0.12172 \end{bmatrix} \text{ dan matriks}$$

distribusi peluang transisi *state*-nya (A) adalah

$$\begin{bmatrix} 0.29627 & 0.29334 & 0.41039 \\ 0.27018 & 0.51084 & 0.21898 \\ 0.25072 & 0.58997 & 0.1593 \end{bmatrix}$$

- (5) Untuk *state* sekarang adalah *sibilant* dengan *state* berikutnya adalah konsonan stop dan konsonan frikatif, maka matriks distribusi peluangnya

$$(\pi) \text{ adalah } \begin{bmatrix} 0.81754 \\ 0.13843 \\ 0.044025 \end{bmatrix} \text{ dan matriks}$$

distribusi peluang transisi *state*-nya (A) adalah

$$\begin{bmatrix} 0.39108 & 0.21404 & 0.39487 \\ 0.37383 & 0.21996 & 0.40621 \\ 0.39173 & 0.28741 & 0.32086 \end{bmatrix}$$

- (6) Untuk *state* sekarang adalah *sibilant* dengan *state* berikutnya adalah konsonan frikatif dan konsonan stop, maka matriks distribusi peluangnya

$$(\pi) \text{ adalah } \begin{bmatrix} 0.90604 \\ 0.04829 \\ 0.04567 \end{bmatrix} \text{ dan matriks}$$

distribusi peluang transisi *state*-nya (A) adalah

$$\begin{bmatrix} 0.14661 & 0.1809 & 0.67249 \\ 0.21332 & 0.65283 & 0.13385 \\ 0.24472 & 0.55067 & 0.20462 \end{bmatrix}$$

Dengan menggunakan model tersebut, proses klasifikasi terlebih dahulu mencari matriks distribusi peluang observasi (B) dan kemudian dengan algoritma Viterbi dicari nilai *likelihood* dari tiap *state*. *State* dengan nilai *likelihood* terbesar menyatakan suatu kondisi memiliki kedekatan dengan *state* tersebut.

Tabel 1 menunjukkan hasil pengujian hasil klasifikasi sinyal suara menggunakan HMM. Pengujian menggunakan 90 sampel suara dengan 30 ucapan untuk tiap fonem yang diklasifikasikan.

Tabel 1. Hasil Uji Klasifikasi Sinyal Suara Menggunakan HMM (dalam %)

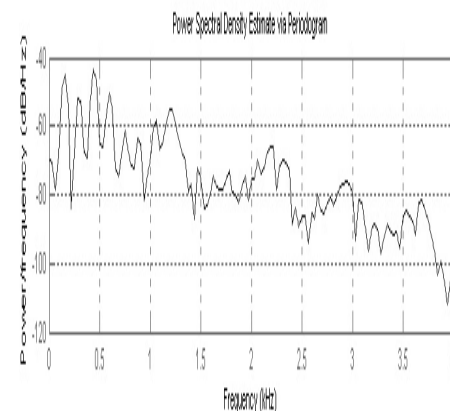
TS \ F	S	KF	KS	TK
S	73,33	10	13,33	3,33
KF	16,67	63,33	16,67	3,33
KS	13,33	10	70	6,67

Keterangan :

- TS : Terdeteksi sebagai
- F : Fonem
- S : *Sibilant*
- KF : Konsonan frikatif
- KS : Konsonan stop
- TK : Tidak terklasifikasi

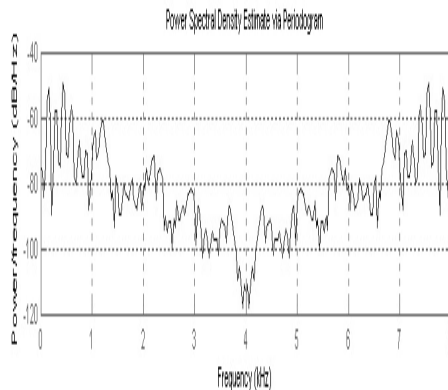
Pada Tabel 1 terlihat bahwa dalam proses klasifikasi sinyal suara pada penelitian ini masih terjadi kesalahan sekitar 30%. Kesalahan klasifikasi sinyal suara berdampak pada kesalahan pemberian respon pada bagian sinyal suara yang akan dimodifikasi.

Contoh hasil ekspansi lebar pita secara artifisial dalam domain frekuensi dapat dilihat secara lengkap dalam Gambar 9.

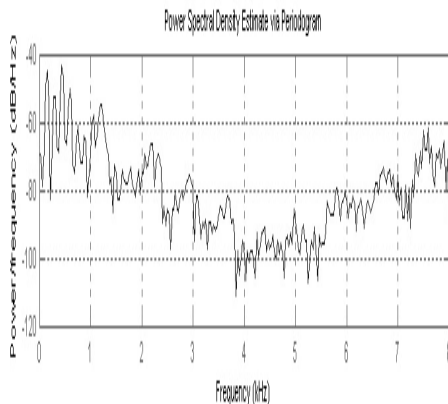


Gambar 9 (a) Spektrum Sebelum Mengalami Proses *Upsampling*

Pada Gambar 9(c) terlihat spektrum pada salah satu *frame* yang telah mengalami proses ekspansi lebar pita. Pada daerah frekuensi tinggi (4kHz – 8 kHz) telah mengalami modifikasi.



Gambar 9 (b) Spektrum Sesudah Mengalami Proses *Upsampling*



Gambar 9 (c) Spektrum Sesudah Mengalami Proses Ekspansi Lebar Pita

Untuk mengetahui kualitas setiap sinyal suara sintesis, dilakukan pengujian suara secara subyektif. Pengujian dilakukan dengan memperdengarkan sinyal suara sintesis kepada beberapa pendengar dan diminta untuk memberikan *Mean Opinion Score* (MOS), dan nilai MOS yang digunakan berkisar dari 1 sampai 5 dan nilai 1 menyatakan kualitas yang sangat kurang dan nilai 5 menyatakan nilai yang sangat baik. Pengujian dilakukan terhadap 20 orang responden, dan hasil pengujian terhadap sinyal suara sintesis dapat dilihat pada Tabel 2.

Pada Tabel 2 terlihat bahwa sinyal suara yang mengalami ekspansi memiliki artikulasi yang cukup jelas dan dapat dimengerti oleh responden walaupun distorsi masih terdengar.

Tabel 2. Nilai MOS pada Tiap *File* Sinyal Suara Sintesis

Nama <i>File</i>	Nilai MOS
bribin_ABE.wav	3,55
Presiden_ABE.wav	3,45
Simón_ABE.wav	3,7
universitas_ABE.wav	3,65
varian_ABE.wav	3,6
Video_ABE.wav	3,55
Zefa_ABE.wav	3,6

KESIMPULAN

Dari hasil pengujian yang telah dilakukan dapat disimpulkan bahwa sinyal suara sintesis yang dihasilkan dari proses ekspansi lebar pita secara artifisial berhasil meningkatkan artikulasi sinyal suara sehingga lebih dapat dimengerti oleh pendengar walaupun distorsi masih terdengar. Hal ini disebabkan oleh proses klasifikasi yang dilakukan belum cukup baik, sehingga mengakibatkan beberapa *frame* menerima respons yang tidak tepat.

Pada penelitian selanjutnya, untuk mengurangi kesalahan klasifikasi, mungkin perlu dikembangkan model HMM yang lebih baik, dengan menggunakan metode ekstraksi fitur yang berbeda seperti *melcepstrum* atau dapat pula digunakan metode pengenalan ucapan lain seperti jaringan syaraf tiruan dan *automatic speech recognition*.

DAFTAR PUSTAKA

- Kallio, L., 2002, *Artificial Bandwidth Expansion of Narrowband Speech in Mobile Communication System Master Thesis*, Helsinki University of Technology.
- Rabiner, L., 1989, *A Tutorial on Hidden Markov Model and Selected Application in Speech Recognition*, *Proceeding of the IEEE*, vol. 77, no. 2.
- Rabiner, L. and Juang, B.H., 1993, *Fundamentals of Speech Recognition*, Prentice Hall, United States.
- Yasukawa, H., 1994, Quality Enhancement of Band Limited Speech by Filtering and Multirate Techniques, *Proceedings of International Conference on Spoken La-*

nguage Processing, pp. 1607-1610.

Yasukawa,H., 1996, Signal Restoration of Broad Band Speech Using Non-linear Processing, *Proceedings of Europe Signal Processing Conference*, pp. 987-990